



## Urban traffic event detection using Twitter data

Das, Rahul Deb ; Purves, Ross S

**Abstract:** Understanding traffic events is important for urban policy making and transport management. Traffic events could be related to traffic congestion, transportation infrastructure issues, parking issues, to name a few. Currently, traffic events are monitored through static sensors e.g., CCTV camera, loop detectors which have limited spatial coverage and high maintenance cost. Thus, we attempt to use the concept of citizens as sensors and develop a cost-effective model to understand urban traffic events from unstructured and informal tweets. So far existing works attempted to classify tweets either in traffic or non-traffic categorization [1], [3], [4]. Most of the state-of-the-art have used geotagged tweets for identifying traffic events [2], which accounted for only 1%-3% total tweet population, and thus lots of useful information in the untagged tweets may be lost. Some other works explored a number of abstract topics related to urban transportation and environment, however without retrieving any spatial information from the tweet [5], [6]. The main contribution of this work is, in contrast to the earlier works, this research explores untagged tweets to detect traffic events and developed a novel framework (Fig. 1, 2) that does not only categorize traffic related tweets but also retrieve locations of the traffic events from the tweet content. The model has been tested in the city of Mumbai in India where people use different local place names which are often informal and hard to detect using a traditional named entity recognition systems. To detect the locations of the traffic events we developed a hybrid georeferencing model that consists of a supervised model and a number of spatial rules that can handle informal place names and vernacular geographical aspects. For tweet categorization we used a binary classifier based on Decision Tree (DT) with 0.65 precision and 0.57 recall. The tweets are manually labelled into either traffic or non-traffic. Then the classifier is trained using a bag-of-words model. In the next phase, a hybrid georeferencing model is developed. The proposed georeferencing model consists of a pre-trained StanfordNER on the top tier and two spatial rule-based layers in the subsequent tiers (Fig. 3). The rules are based on spatial prepositions, object types and vernacular place names in India. Out of 1143 annotated place names the model can correctly retrieve 691 place names. To disambiguate and geocode the place names, OpenStreetMap has been used. This work shows Twitter can be useful for detecting urban events in Mumbai. One of the challenges in georeferencing the traffic event location lies in the way people mention the place names. The same place name may be mentioned differently by different people or it may not be present in the gazetteer, e.g., OpenStreetMap, which causes difficulty in toponym recognition and disambiguation. In this work the toponyms (retrieved from tweet content) are mapped to precise geo-coordinates to indicate traffic locations. However, traffic events can stretch along a street segment or over a region. Future work will look into understanding the spatial extent of the affected area from other contextual cues and spatial relationships retrieved from the tweet content.

Originally published at:

Das, Rahul Deb; Purves, Ross S (2018). Urban traffic event detection using Twitter data. In: VGI Geovisual Analytics Workshop, colocated with BDVA 2018, Konstanz, 19 October 2018, SNSF.

# Urban Traffic Event Detection using Twitter Data

Rahul Deb Das  
Department of Geography  
University of Zurich  
Switzerland  
rahul.das@geo.uzh.ch

Ross S. Purves  
Department of Geography  
University of Zurich  
Switzerland  
ross.purves@geo.uzh.ch

**Abstract**—Understanding traffic events is important for urban policy making and transport management. Traffic events could be related to traffic congestion, transportation infrastructure issues, parking issues, to name a few. Currently, traffic events are monitored through static sensors e.g., CCTV camera, loop detectors which have limited spatial coverage and high maintenance cost. Thus, we attempt to use the concept of citizens as sensors and develop a cost-effective model to understand urban traffic events from unstructured and informal tweets. So far existing works attempted to classify tweets either in traffic or non-traffic categorization [1], [3], [4]. Most of the state-of-the-art have used geotagged tweets for identifying traffic events [2], which accounted for only 1%-3% total tweet population, and thus lots of useful information in the untagged tweets may be lost. Some other works explored a number of abstract topics related to urban transportation and environment, however without retrieving any spatial information from the tweet [5], [6].

The main contribution of this work is, in contrast to the earlier works, this research explores untagged tweets to detect traffic events and developed a novel framework (Fig. 1, 2) that does not only categorize traffic related tweets but also retrieve locations of the traffic events from the tweet content. The model has been tested in the city of Mumbai in India where people use different local place names which are often informal and hard to detect using a traditional named entity recognition systems. To detect the locations of the traffic events we developed a hybrid georeferencing model that consists of a supervised model and a number of spatial rules that can handle informal place names and vernacular geographical aspects.

For tweet categorization we used a binary classifier based on Decision Tree (DT) with 0.65 precision and 0.57 recall. The tweets are manually labelled into either traffic or non-traffic. Then the classifier is trained using a bag-of-words model. In the next phase, a hybrid georeferencing model is developed. The proposed georeferencing model consists of a pre-trained StanfordNER on the top tier and two spatial rule-based layers in the subsequent tiers (Fig. 3). The rules are based on spatial prepositions, object types and vernacular place names in India. Out of 1143 annotated place names the model can correctly retrieve 691 place names. To disambiguate and geocode the place names, OpenStreetMap has been used. This work shows Twitter can be useful for detecting urban events in Mumbai. One of the challenges in georeferencing the traffic event location lies in the way people mention the place names. The same place name may be mentioned differently by different people or it may not be present in the gazetteer, e.g., OpenStreetMap, which causes difficulty in toponym recognition and disambiguation. In this work the toponyms (retrieved from tweet content) are mapped to precise geo-coordinates to indicate

traffic locations. However, traffic events can stretch along a street segment or over a region. Future work will look into understanding the spatial extent of the affected area from other contextual cues and spatial relationships retrieved from the tweet content.

**Index Terms**—traffic, tweet, georeference, toponym, StanfordNER

## I. ACKNOWLEDGMENT

We would like to acknowledge the Swiss National Science Foundation (SNSF) for their support in his research. We would also like to thank the reviewers for their valuable and insightful comments.

## REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "Real-Time Detection of Traffic From Twitter Stream Analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16(4), pp. 2269–2283, 2015.
- [2] D. A. Kurniawan, S. Wibirama, and N. A. Setiawan, "Real-time traffic classification with Twitter data mining," in 8th International Conference on Information Technology and Electrical Engineering (ICITEE), 2016.
- [3] S. Klaithin, and C. Haruechaiyasak, "Traffic information extraction and classification from Thai Twitter," in 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2016.
- [4] A. Salas, P. Georgakis, and Y. Petalas, "Incident detection using data from social media," in *IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017.
- [5] Y. Zhou, S. De, and K. Moessner, "Real World City Event Extraction from Twitter Data Streams," in *Procedia Computer Science*, vol. 98, pp. 443–448, 2016.
- [6] A. F. Hidayatullah, and M. R. Ma'arif, "Road traffic topic modeling on Twitter using latent dirichlet allocation," in *International Conference on Sustainable Information Engineering and Technology (SIET)*, 2017.

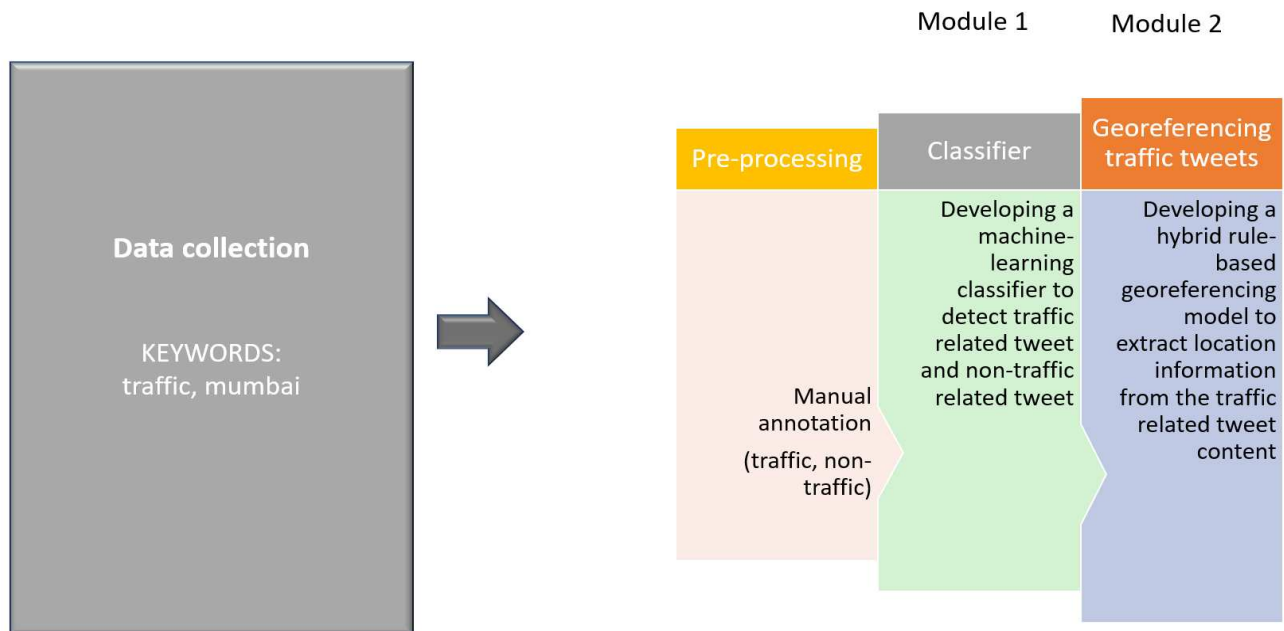


Fig. 1. Workflow.

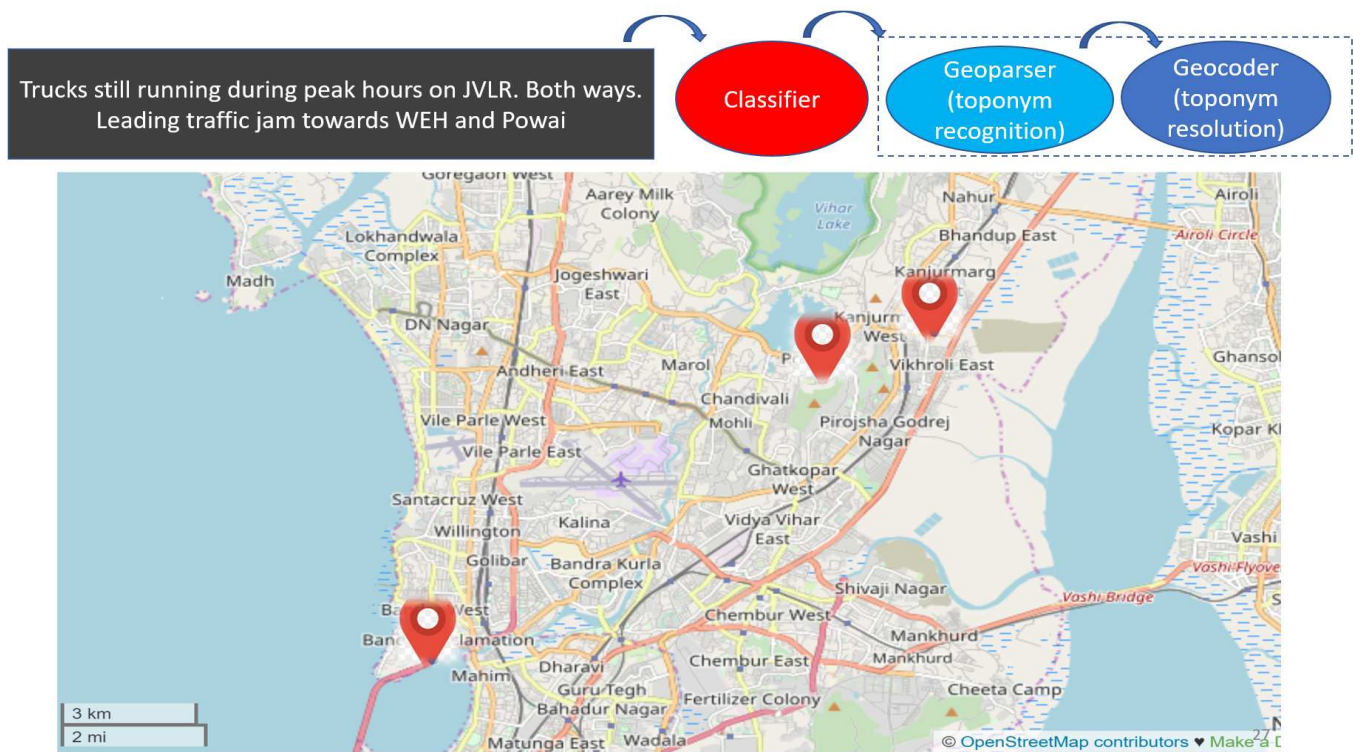


Fig. 2. Detecting traffic events and their locations through multi-layered georeferencing model.

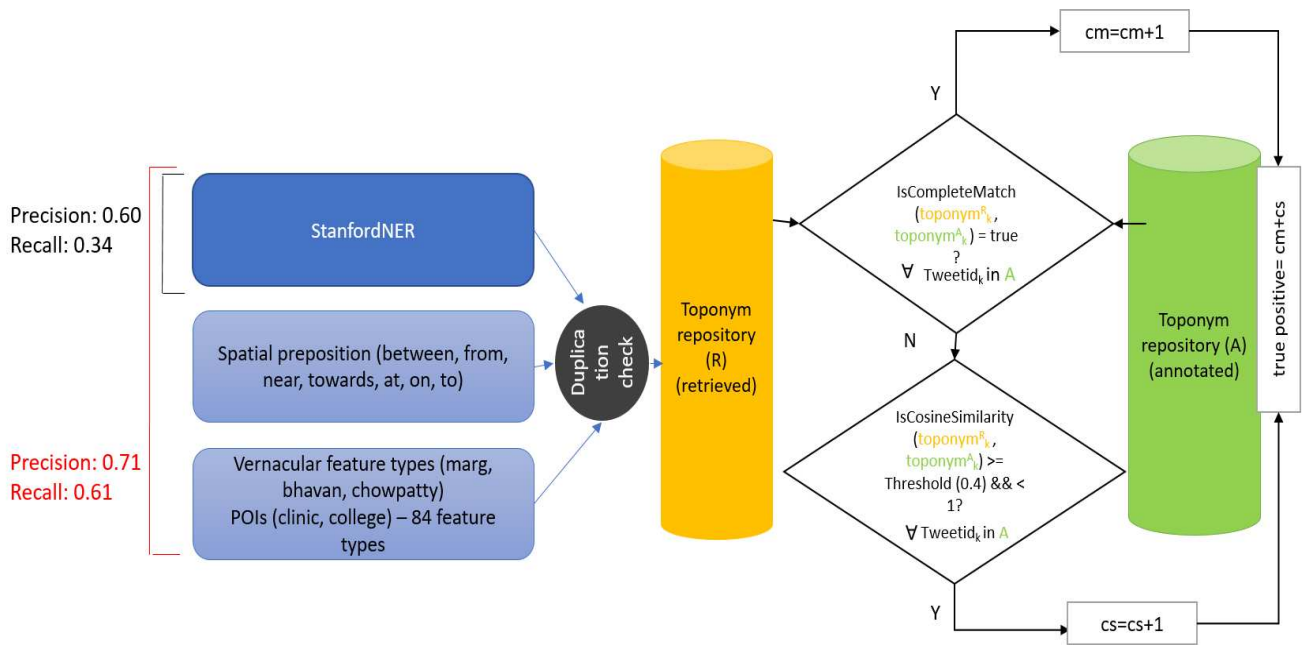


Fig. 3. The proposed multi-layered hybrid georeferencing model.